Iterative Estimation Correcting for Error Autocorrelation in Short Panels

Rembert De Blander*

October 12, 2007

Abstract

This paper presents an iterative estimation procedure that estimates and corrects for serial correlation of the disturbances in short panels. Controlling for error autocorrelation is a prerequisite for consistent estimation in models with lagged dependent variables. In addition it allows to discern between different behavioural mechanisms underlying state persistence. The basic philosophy of iterative estimation is to assume some information on the basis of which the parameters of the postulated structural model are easily estimated. These estimates subsequently allow to update the assumed information and the complete cycle is repeated until convergence. The unobserved or latent variables considered here are the residuals from previous periods.

While the main result is valid for models that allow for the explicit calculation of Cox and Snell's (1968) generally defined residuals, which in turn are allowed to exhibit a very general temporal dependence structure, attention is subsequently restricted to AR correlated disturbances, since an MA error process would require strong assumptions on the initial values for consistency when $N \to \infty$, with T fixed. The method is finally applied to short panel data models with fixed effects and lagged dependent variables as well.

Keywords: Serial correlation, panel data, iterative estimation, true state dependence

JEL classification: C23

1 Introduction

Controlling for error autocorrelation is a prerequisite for consistent estimation in models with lagged dependent variables. In addition it allows to discern between dif-

*H.I.V.A. and C.E.S., K.U.Leuven and K.U.Brussel, Parkstraat 47, B-3000 Leuven, Belgium; *E-mail*: Rembert@DeBlander.eu

ferent behavioural mechanisms underlying state persistence. Indeed, observed state persistence can either be driven by persistence in the states determinants (whether observed or unobserved), or it can be caused by an alteration of the persons characteristics. The latter situation is termed *true state dependence* in the econometrics literature, while the former dependence is called *spurious* (Heckman (1978a)). The same issue is called the distinction between *true* and *apparent* contagion in the biostatistics literature (Aitkin and Alfò (2003)).

Unobserved individual-specific characteristics are an extreme form of error persistence, and thus, when not controlled for, an important source of spurious state dependence. In addition, personal characteristics that remain relatively stable, like intelligence or ambition, are suspect of being correlated with almost every economically interesting phenomenon and, thus, when they remain unobserved, a potential source of bias, both in static and in dynamic models. Not surprisingly, large efforts in the dynamic panel data literature are devoted to the development of estimators that accommodate these individual effects.

When we are willing to postulate an error distribution, any kind of error dependence can be modelled using the method of maximum likelihood (*ML*). Considering an error components model, *ML* estimation is suggested by Bhargava and Sargan (1983), who treat the individual effects as random, and by Hsiao *et al.* (2002), treating the individual effects as fixed.

Without having to specify an error distribution, Liang and Zeger (1986) estimate the parameters of interest via generalized estimating equations[1] (*GEE*) and the nuisance parameters modeling the error dependence by moment estimates in terms of the residuals. Chaganty (1997) adapts the *GEE* by improving the estimation of the nuisance parameters. Hansens (1982) generalized method of moments (*GMM*) estimator in conjunction with predetermined instruments, permits the disturbances to be serially correlated. Keane and Runkle (1992) adapt Hayashi and Sims (1983) forward filtered estimator (*FFE*) to panel data. This estimator eliminates serial correlation by linearly combining the observations for period t and later, thus preserving the models orthogonality conditions with respect to predetermined instruments. While this estimator can be applied to first-differenced fixed-effects models, it relies on a consistent estimate of the variance matrix of the serially correlated disturbances, making it unsuitable when lagged dependent variables are included as regressors.

Ahn and Schmidt (1995, 1997), Anderson and Hsiao (1981), Arellano and Bond (1991), Arellano and Bover (1995) all explicitly consider the lagged dependent vari-

[1] The *GEE* are the quasi score functions of the quasi-likelihood.

able error component model

$$y_{it} = \alpha y_{i,t-1} + \beta' x_{it} + u_{it} \tag{1}$$

$$u_{it} = \mu_i + \varepsilon_{it} \tag{2}$$

While Anderson and Hsiao (1981) suggested estimating the first-differenced model using either $y_{i,t-2}$ or $\triangle y_{i,t-2}$ as instruments for $\triangle y_{i,t-1}$, Arellano and Bond (1991) remark that the former choice of instrument is merely a subset of the vector all feasible instruments $(y_{i,t-2}, \ldots, y_{i1})$. Arellano and Bover (1995), on the other hand, recommend estimating the model in levels using the instruments $(\triangle y_{i,t-2}, \ldots, \triangle y_{i1})$ for $y_{i,t-1}$. While Ahn and Schmidt (1995, 1997) improve the efficiency of these *GMM* estimators by considering additional moment conditions reflecting covariance restrictions and initial conditions, Doran and Schmidt (2006) improve their finite sample properties by using principal components of the weighting matrix.

The error component specification (2) can be extended to either

$$u_{it} = \mu_i + \rho u_{i,t-1} + \varepsilon_{it}, \tag{3}$$

or

$$u_{it} = \mu_i + \rho \varepsilon_{i,t-1} + \varepsilon_{it}. \tag{4}$$

Lillard and Willis (1978) consider (1) with $\alpha \equiv 0$ and with error specification (3), treating μ_i as random effects. Bhargava *et al.* (1982) adapt the Durbin-Watson (DW) and the Berenblut-Webb statistics to panel data with fixed effects by basing them on *LSDV* residuals. On the basis of the DW statistic, they construct an estimator of the $AR(1)$ error serial correlation. However, this approach relies on the $N \to \infty$ consistency of the *LSDV* estimator and is thus not applicable in the presence of lagged dependent variables. Baltagi and Li (1995) present different LM tests to discern (2), (3) and (4), both when μ_i are random or fixed. Arellano and Bonds (1991) *GMM* estimator retains its consistency in the presence of fixed effects and $MA(q)$ errors when $\triangle y_{it}$ is instrumented by $(y_{i,t-q-1}, \ldots, y_{i1})$. What seems missing in the literature, is an estimator for fixed effects and $AR(p)$ errors for the pure lagged dependent variable model, i.e. model (1) with $\beta \equiv 0$, without relying on strictly exogenous instruments.

Indeed, none of the above estimators is trivially modifiable such that it is consistent under the error specification (3) without making extra assumptions on the exogeneity of the x_{it}, since $\mathrm{E}[y_{is} u_{it}] \neq 0$, for all s, t, and thus lagged values of y_{it}

can not be used as instruments. One feasible option out of this stalemate is the recourse to iterative estimation, the basic philosophy of which is to assume some information on the basis of which the parameters of the postulated structural model are easily estimated. These estimates subsequently allow to update the assumed information and the complete cycle is repeated until convergence. The best-known iterative algorithm is the expectation-maximization (*EM*) algorithm (Dempster *et al.* (1977), Hartley (1958)), which accommodates incomplete data in a *ML* context. In the context of this paper, the latent variables are either the lagged residuals $u_{i,t-j}$ or the lagged innovations $\varepsilon_{i,t-j}$. The iterative approaches of Dominitz and Sherman (2005) and of Pastorello et al. (2003) both encompass the *EM* algorithm and provide us with an asymptotic theory.

In the next section, the general principle is formulated, which is subsequently applied to lagged dependent variable models with autoregressive idiosyncratic error components in section 3. Section 4 presents some Monte Carlo results, while section 5 concludes. Proofs of consistency and asymptotic normality are relegated to the appendix.

2 Iterative Estimation

2.1 General principle

Consider the general model

$$y_{it} = g\left(x_{it}, u_{it}; \beta_0\right), \tag{5}$$

where $i = 1, \ldots, N$, $t = 2, \ldots, T$, $g(\cdot)$ is a known function, y_{it} is an observed outcome, x_{it} is a vector of observed determinants, u_{it} a scalar error term and β_0 the true parameter vector. Suppose now that the equation

$$y_{it} = g\left(\hat{u}_{it}; x_{it}, \hat{\beta}\right)$$

has a unique solution for $\hat{u}_{it}$, i.e.

$$\hat{u}_{it} = g^{(-1)}\left(y_{it}; x_{it}, \hat{\beta}\right), \tag{6}$$

which is Cox and Snell's (1968) general definition of a residual. Evidently, it holds that $u_{it} = g^{(-1)}\left(y_{it}; x_{it}, \beta_0\right)$. Suppose further that there is some form of temporal dependency between the disturbances, such that u_{it} is a known function $h(\cdot)$ of previous disturbances, contemporaneous innovation ε_{it}, and previous innovations,

$$u_{it} = h\left(\varepsilon_{it}; u_{i,t-1}, u_{i,t-2}, \ldots, u_{i,t-p}, \varepsilon_{i,t-1}, \varepsilon_{i,t-2}, \ldots, \varepsilon_{i,t-q}, \rho_0\right) \qquad (7)$$

that has a unique solution for ε_{it}, i.e.

$$\varepsilon_{it} = h^{(-1)}\left(u_{it}; u_{i,t-1}, u_{i,t-2}, \ldots, u_{i,t-p}, \varepsilon_{i,t-1}, \varepsilon_{i,t-2}, \ldots, \varepsilon_{i,t-q}, \rho_0\right).$$

The innovation ε_{it} can thus be expressed in function of observed contemporaneous information, lagged disturbances and/or innovations and a finite vector of parameters $\theta_0 = (\beta_0', \rho_0')'$, as

$$\begin{aligned}
\varepsilon_{it} &= h^{(-1)}\left(g^{(-1)}\left(y_{it}; x_{it}, \beta_0\right); u_{i,t-1}, u_{i,t-2}, \ldots, u_{i,t-p}, \varepsilon_{i,t-1}, \varepsilon_{i,t-2}, \ldots, \varepsilon_{i,t-q}, \rho_0\right) \\
&= \tilde{h}^{(-1)}\left(y_{it}; x_{it}, u_{i,t-1}, u_{i,t-2}, \ldots, u_{i,t-p}, \varepsilon_{i,t-1}, \varepsilon_{i,t-2}, \ldots, \varepsilon_{i,t-q}, \theta_0\right).
\end{aligned}$$

Two common error structures are the autoregressive $AR(p)$

$$u_{it} = \varepsilon_{it} + \sum_{s=1}^{p} \rho_{0;s} u_{i,t-s}$$

and the moving average $MA(q)$ type

$$u_{it} = \varepsilon_{it} + \sum_{s=1}^{q} \rho_{0;s} \varepsilon_{i,t-s}.$$

Consider now the hypothetical situation where the past disturbances u_{is} and innovations ε_{is}, $s < t$, are known and suppose that following assumption holds[2] .

Assumption 1. Given that u_{is} and ε_{is}, $1 < s < t$, are observed, there exists an estimator $\hat{\theta}_N = \left(\hat{\beta}_N', \hat{\rho}_N'\right)'$ for which

1. $\underset{N\to\infty}{\text{plim}} \hat{\theta}_N = \theta_0$,

2. $\sqrt{N}\left(\hat{\theta}_N - \theta_0\right) \overset{d}{\longrightarrow} N(0; \Sigma_{\hat{\theta}})$.

I thus explicitly consider the panel data case with finite time dimension and a large number of observational units, a configuration prevalent in many micro-economic

[2]For the more general case of N^δ-consistency, see Dominitz and Sherman (2005). The extension to asymptotic bias-corrected $\hat{\theta}$ will not be undertaken in the context of this paper.

panels. The innovations ε_{it} can then be estimated consistently by

$$\hat{\varepsilon}_{it} = \tilde{h}^{(-1)}\left(y_{it}; x_{it}, u_{i,t-1}, u_{i,t-2}, \ldots, u_{i,t-p}, \varepsilon_{i,t-1}, \varepsilon_{i,t-2}, \ldots, \varepsilon_{i,t-q}, \hat{\theta}_N\right). \quad (8)$$

The basic idea is now to iterate the following algorithm.

Algorithm 1 Iterative estimation

Use some initial estimate $\hat{\theta}_N^{(0)}$ to predict $\hat{\varepsilon}_{it}^{(0)}$ and $\hat{u}_{it}^{(0)}$.
Iterate until convergence:

1. Given $\hat{\varepsilon}_{it}^{(k)}$ and $\hat{u}_{it}^{(k)}$, execute the chosen estimator and obtain $\hat{\theta}_N^{(k+1)}$.
2. Predict $\hat{\varepsilon}_{it}^{(k+1)}$ and $\hat{u}_{it}^{(k+1)}$, using $\hat{\theta}_N^{(k+1)}$.

Starting from some initial estimate $\hat{\theta}_N^{(0)}$, iteration of both steps ad infinitum results in the series of estimates $\hat{\theta}_N^{(1)}, \hat{\theta}_N^{(2)}, \ldots, \hat{\theta}_N^{(\infty)}$. It is this last estimate I propose to use as an estimator for θ_0.

While Assumption 1 includes assumptions on the initial conditions neccessary for consistency and asymptotic normality, it is worthwile to discuss them here explicitly. A first way to treat initial conditions is to assume that $u_{is} = \varepsilon_{is} = 0$, $s \leq 0$. All cross-sections can be used in both the estimation step and the prediction step of algorithm 9. When $q = 0$, i.e. no past innovations are explicitly present in $h(\cdot)$, it is possible to discard any assumption on the disturbances prior to the window of observation $1 \leq t \leq T$, by treating the p initial cross-sections different from the later ones. Only the last $T - p$ cross-sections are used to estimate the parameter vector θ_0, and the disturbances are estimated by (6). The first p cross-sections are thus only used to predict the disturbances $u_{i,s}$, $1 \leq s \leq p$.

In case of a model with additive errors various other options are possible. Consider for instance an $AR(p)$ error structure with initial conditions $u_{is} = 0$, $s \leq 0$, then the treatment of the initial conditions can still be refined. In the first cross-section, only β_0 is estimated and the disturbances u_{i1}, are obtained by (6). The r^{th} cross-section ($r \leq p$) is used in the estimation of β_0 as well as the first $r - 1$ components of ρ_0, since $u_{i1}, \ldots, u_{i,r-1}$ can be estimated. The remaining error terms in this cross-section are given by $v_{ir} = \varepsilon_{ir} + \sum_{s=r}^{p} \rho_{0;s} u_{i,r-s}$. From cross-section $p + 1$ onwards, the full vector θ_0 is estimated and the innovations are estimated by (8). While such a cross-section dependent error definition will result in cumbersome likelihood functions, it is easily accommodated in a regression framework.

2.2 Asymptotic properties

In order to study the asymptotic properties of $\hat{\theta}_N^{(\infty)}$, I define the equations that identify the unfeasible $\hat{\theta}_N$ as $U_N(\theta;\psi) = N^{-1}\sum_{i,t} U_{it}(\theta;\psi) = 0$ and their probability limit for $N \to \infty$ as $U(\theta;\psi) = 0$. These identifying equations or estimating functions (McCullagh and Nelder (1999)) are here not only functions of the parameters of interest, θ, but of some nuisance parameters ψ as well. The latter are not estimated, but are present in the identifying equations but are present in these equations through their influence on the latent variables, i.e. the unobserveds u_{is}, $s < t$. In the context of *ML* estimation U_{NT} are the likelihood (or efficient score) equations (Cox and Hinkley (1982, p.283)), while in the context of *GMM* estimation, the U_N are the weighted sample moment restrictions[3] (Hansen (1982)). Assuming that $\hat{\theta}_N$ consistently estimates θ_0 for $N \to \infty$ when u_{is}, $s < t$, would have been observed, is tantamount to saying that the solution $\hat{\theta}_N$ of $U_N(\theta;\psi_0)$ converges in probability to the solution of $U(\theta;\psi_0)$, which is θ_0 by assumption. All assumptions for consistency and asymptotic normality of the considered unfeasible estimator are implicitly made in Assumption 1. In addition, it is assumed that $\varepsilon_{it} \sim IID(0, \sigma_\varepsilon^2)$.

To the first order, $U_N(\theta;\psi)$ is given by

$$\begin{aligned} U_N\left(\hat{\theta}_N;\psi\right) &= U_N(\theta_0;\psi_0) + \left.\frac{\partial U_N(\theta;\psi)}{\partial \theta'}\right|_{\substack{\theta=\theta_0 \\ \psi=\psi_0}} \left(\hat{\theta}_N - \theta_0\right) \\ &\quad + \left.\frac{\partial U_N(\theta;\psi)}{\partial \psi'}\right|_{\substack{\theta=\theta_0 \\ \psi=\psi_0}} (\psi - \psi_0) + o_p\left(N^{-\frac{1}{2}}\right), \end{aligned} \tag{9}$$

which, after taking probability limits, results in

$$\left(\hat{\theta} - \theta_0\right) = M(\psi - \psi_0), \tag{10}$$

with

$$M = -J^{-1}K, \tag{11}$$

$$J = \left.\frac{\partial U(\theta;\psi)}{\partial \theta'}\right|_{\substack{\theta=\theta_0 \\ \psi=\psi_0}}. \tag{12}$$

[3] An expression is provided in section 3.

and

$$K = \left.\frac{\partial U(\theta;\psi)}{\partial \psi'}\right|_{\substack{\theta=\theta_0\\ \psi=\psi_0}} \left.\frac{\partial U(\theta;\psi)}{\partial (u',\varepsilon')}\right|_{\substack{\theta=\theta_0\\ \psi=\psi_0}} \left.\frac{\partial (u',\varepsilon')'}{\partial \psi'}\right|_{\psi=\psi_0}. \tag{13}$$

Define M_N, respectively J_N as the sample analogs of these matrices.

In the context considered here, the vector ψ represents the parameter estimate $\hat{\theta}^{(j-1)}$ from the previous iteration and $\hat{\theta}$ in (10) gives the new parameter estimate $\hat{\theta}^{(j)}$. A sufficient condition for the asymptotic bias of this new estimate to be smaller than the asymptotic bias of the previous estimate can be stated in terms of the spectral radius of M.

Definition 1. The *spectral radius* $\tau(A)$ of a square matrix A is defined by $\tau(A) = \max\{|\lambda| : \lambda \in \sigma(A)\}$, where $\sigma(A)$ denotes the spectrum, or set of all eigenvalues λ of A (Horn and Johnson (1994)).

Assumption 2 (consistency). For the matrix M as defined in (11), it holds that $\tau(M) < 1$.

For any vector x and any square matrix A, we have that $\|y\| = \|Ax\| \leq \tau(A)\|x\|$. Under Assumption 2 it thus holds that $\left\|\hat{\theta}^{(j)} - \theta_0\right\| \leq \left\|\hat{\theta}^{(j-1)} - \theta_0\right\|$ and the mapping $\hat{\theta}^{(j-1)} \rightarrow \hat{\theta}^{(j)}$ is called a *contraction mapping* (Pastorello et al. (2003)). For the iteration to converge in finite samples as well, we have to impose the same restriction on M_N, the sample analog of M.

Premultiplication of (9) by $\sqrt{N}$, and iteratively applying the considered estimator, results in

$$\sqrt{N}\left(\hat{\theta}_N^{(k)} - \theta_0\right) = -J_N^{-1}\sqrt{N}U_N(\theta_0;\psi_0) + M_N^k\sqrt{N}\left(\hat{\theta}_N^{(0)} - \theta_0\right) + o_p(1).$$

For $\sqrt{N}\left(\hat{\theta}_N^{(\infty)} - \theta_0\right)$ to converge to a zero mean normal variate independent of the initial estimator $\hat{\theta}_N^{(0)}$, we need to impose that the second term above is also $o_p(1)$, which holds under following Assumption.

Assumption 3 (asymptotic normality). The number of iterations k satisfies the inequality[4]

$$k(N) \quad > \quad \frac{-\ln N}{2\ln\tau(M)}. \tag{14}$$

When Assumption 2 holds, the inconsistency of the initial estimate $\hat{\theta}_N^{(0)}$ becomes negligibly small at a fast enough rate to have no impact on the asymptotic distribution of $\hat{\theta}_N^{(\infty)}$.

The variance of $\sqrt{N}\left(\hat{\theta}_N^{(\infty)} - \theta_0\right)$ is derived by noticing that $u_{i,t-1}, \ldots, u_{i,t-p}$, $\varepsilon_{i,t-1}, \ldots, \varepsilon_{i,t-q}$ are not observed, but are estimated. The variance of estimators obtained by using synthetic regressors are well-known (see Murphy and Topel (1985), Parke (1986), Pierce (1982) and Randles (1982)). Reorganization of (9) premultiplied by $\sqrt{N}$, considered after infinitely many iterations results in

$$\sqrt{N}\left(\hat{\theta}_N^{(\infty)} - \theta_0\right) \quad = \quad -\left(I - M_N\right)^{-1} J_N^{-1}\sqrt{N}U_N\left(\theta_0; \psi_0\right) + o_p\left(1\right),$$

which, together with the Lyapunov *CLT* proves following Theorem.

Theorem 1. Under the assumption that (14) holds, the iterative estimator $\hat{\theta}_N^{(k)}$ based on the unfeasible estimator $\hat{\theta}_N$ for which Assumption 1 holds has an asymptotic distribution given by

$$\sqrt{N}\left(\hat{\theta}_N^{(\infty)} - \theta_0\right) \overset{d}{\longrightarrow} N\left(0; \left(I - M\right)^{-1}\Sigma_{\hat{\theta}}\left(I - M\right)^{-1}\right),$$

where M is defined in (11).

Theorem 1 is follows immediately from Pastorello et al. (2003) and is consistent with Oakes (1999).

Example 2. Consider a linear model with $AR(1)$ disturbances, i.e. a structural model of the form

$$y_{it} \quad = \quad \alpha_0' x_{it} + \rho_0 u_{i,t-1} + \varepsilon_{it}$$

and the sample identifying equations of the unfeasible *OLS*-estimation of y_{it} on x_{it} and $u_{i,t-1}$ are given by

$$U_N \quad = \quad \sum_{it}\left(y_{it} - \alpha_0' x_{it} - \rho_0 u_{i,t-1}\right)\begin{pmatrix} x_{it} \\ u_{i,t-1} \end{pmatrix} = 0.$$

---

[4] In the more general case that the unfeasible $\hat{\theta}$ is N^{δ}-consistent the equivalent of (14) becomes $k(N) > \delta \ln N / \ln\tau(M)$ (Dominitz and Sherman (2005)).

In this context, $\left.\frac{\partial U_N(\theta;\psi)}{\partial \theta'}\right|_{\substack{\theta=\theta_0\\ \psi=\psi_0}}$ is nothing more than the matrix of cross-products of (unfeasible) regressors $(x_{it}', u_{i,t-1})$. Now, we have that

$$\left.\frac{\partial U_N(\theta;\psi)}{\partial \psi'}\right|_{\substack{\theta=\theta_0\\ \psi=\psi_0}} = -\rho_0 \sum_{it} \begin{pmatrix} x_{it} \\ u_{i,t-1} \end{pmatrix} \frac{\partial u_{i,t-1}}{\partial \psi'} + \sum_{it} \varepsilon_{it} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \frac{\partial u_{i,t-1}}{\partial \psi'}$$

with

$$\frac{\partial u_{i,t-1}}{\partial \psi'} = -\begin{pmatrix} x_{i,t-1}' & 0 \end{pmatrix},$$

since $u_{it} = y_{it} - \alpha_0' x_{it}$. Consequently, it holds that

$$\begin{aligned} M &= -\rho_0 \begin{pmatrix} \mathrm{E}\left[x_{it}x_{it}'\right] & 0 \\ 0 & \frac{\sigma_\varepsilon^2}{1-\rho_0^2} \end{pmatrix}^{-1} \begin{pmatrix} \mathrm{E}\left[x_{it}x_{i,t-1}'\right] & 0 \\ 0 & 0 \end{pmatrix} \\ &= -\rho_0 \begin{pmatrix} \Pi_0 & 0 \\ 0 & 0 \end{pmatrix}, \end{aligned}$$

with $\Pi_0 = \left(\mathrm{E}\left[x_{it}x_{it}'\right]\right)^{-1} \mathrm{E}\left[x_{it}x_{i,t-1}'\right]$ and where I have assumed that the $AR(1)$ process defining the disturbances started at $t = -\infty$. Sufficient conditions for $\operatorname*{plim}_{N\to\infty} \hat{\theta}_N^{(\infty)} \to \theta_0$ are that the x_{it} are stationary and that $|\rho_0| < 1$.
The proposed estimator $\hat{\theta}_N^{(\infty)}$ has the advantage that it only requires

$$\mathrm{E}\left[x_{it}\varepsilon_{it}\right] = 0,$$

i.e. it only requires contemporaneous uncorrelatedness. If we simply ignore the first cross-section during estimation, no other conditions are required. Other assumptions leading to consistency are $u_{is} = 0$, $s \leq 0$. The feasible *OLS*-estimation of y_{it} on x_{it} requires the much stronger assumption that the regressors are strictly exogenous, i.e. $\mathrm{E}\left[x_{is}\varepsilon_{it}\right] = 0$, $\forall s, t$.

Remark 1. Convergence of the iterative estimator is only proved locally, i.e. close to θ_0. If the likelihood has multiple local maxima, the parameter space can be divided into domains of convergence, one per local maximum (Arslan et al. 1993)).

In the next section, the proposed iterative procedure will be applied to panel data models with lagged dependent variables with and without fixed effects.

3 Lagged dependent variables

3.1 No fixed effects

The simplest autoregressive linear model is the specification without exogenous covariates

$$y_{it} - \mu_y = \alpha\left(y_{i,t-1} - \mu_y\right) + u_{it}, \qquad |\alpha| < 1, \tag{15}$$

where $i = 1, \ldots, N$, $t = 2, \ldots, T$ and the disturbances follow the $AR\left(1\right)$ process

$$u_{it} = \rho u_{i,t-1} + \varepsilon_{it}, \qquad |\rho| < 1, \tag{16}$$

with $\left(\varepsilon_{i2}, \ldots, \varepsilon_{iT}\right) \sim IID\left(0; \sigma_\varepsilon^2 I_{T-1}\right)$. The variable $y_{i,t-1}$ is clearly correlated with the disturbance u_{it}, which results in biased parameter estimates when the serial correlation is neglected. Remark, however that

$$\mathrm{E}\left[y_{it} - \mu_y \mid y_{i,t-1}, u_{i,t-1}\right] = \alpha\left(y_{i,t-1} - \mu_y\right) + \rho u_{i,t-1},$$

which suggests that application of *OLS*, *ML*, *GMM*, ... on model (15)-(16) would result in a consistent estimator of $\left(\alpha, \rho\right)'$, if $u_{i,t-1}$ would have been observed. Again, given some initial estimate, algorithm 1 provides an estimator.

Example 3. The iterative estimator based on the infeasible *OLS* estimator

$$\left(\sum_{i=1}^N \sum_{t=t_0}^T z_{it} z_{it}'\right)^{-1} \left(\sum_{i=1}^N \sum_{t=t_0}^{T-1} z_{it} y_{it}\right), \tag{17}$$

where $z_{it} = \left(1, y_{i,t-1}, u_{it}\right)$, would be consistent for $N \to \infty$, when $t_0 = 2$, under the condition that $u_{i1} = 0$. However, for $t_0 = 3$ no conditions on the disturbances are needed, since period $t = 2$ is not used in the estimation step but only to predict u_{i1}.

Remark 3. The parameters of structural model (15)-(16) can be estimated consistently by *OLS* estimation of the reduced form

$$y_{it} = \pi_0 + \pi_1 y_{i,t-1} + \pi_2 y_{i,t-j-1} + \varepsilon_{it}, \tag{18}$$

where $\pi_0 = (1-\alpha)(1-\rho)\mu_y$, $\pi_1 = \alpha + \rho$ and $\pi_2 = -\alpha\rho$. A disadvantage of this procedure is that $\sqrt{N}(\hat{\alpha} - \alpha_0, \hat{\rho} - \rho_0)'$, with $\hat{\alpha}, \hat{\rho} = \left(\hat{\pi}_1 \pm \sqrt{\hat{\pi}_1^2 + 4\hat{\pi}_2}\right)/2$ is non-normally distributed. Furthermore it is unclear which root corresponds to $\hat{\alpha}$ and which to $\hat{\rho}$.

3.2 Fixed effects

In the presence of individual-specific intercepts

$$\begin{aligned} y_{it} &= \mu_i + \alpha y_{i,t-1} + u_{it}, \qquad |\alpha| < 1 && (19) \\ u_{it} &= \rho u_{i,t-1} + \varepsilon_{it}, \qquad |\rho| < 1, && (20) \end{aligned}$$

I suggest to consider model (19)-(20) in first-differences

$$\triangle y_{it} = \theta \triangle w_{i,t-1} + \triangle \varepsilon_{it}, \qquad t = 2, \ldots, T,$$

with $w_{it} = (y_{it}, u_{it})'$, and to estimate $\theta = (\alpha, \rho)'$ using the modified Arrelano-Bond (1991) instruments $Z_{it} = \left(z_{i1}', z_{i2}', \ldots, z_{i,t-2}'\right)'$, with $z_{it} = (y_{it}, \triangle u_{it})'$, $t > 1$ and $z_{i1} = y_{i1}$. Define now the matrix of stacked observations $X = (X_1', \ldots, X_i', \ldots, X_N')'$, where the observations for person i are stacked as $X_i = (x_{i3}', \ldots, x_{it}', \ldots, x_{iT}')'$ for all variables except the instruments. The set of instruments for person i are stacked as

$$\begin{aligned} Z_i &= \operatorname{diag}\left[Z_{i1}', \ldots, Z_{i,T-2}'\right] \qquad ((T-2) \times L) \\ &= \begin{pmatrix} z_{i1} & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & z_{i1} & z_{i2} & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & z_{i1} & \cdots & z_{i,T-3} & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & z_{i1} & \cdots & z_{i,T-2} \end{pmatrix} \end{aligned}$$

with $L = (T-2)(T-1)/2$. We can now write the modified Arrelano-Bond (1991) *GMM* estimator as

$$\begin{aligned} \hat{\theta}_{ABn} &= \arg\min_{\theta} N^{-1} (\triangle E' Z) A_N (Z' \triangle E) \\ &= J_N^{-1} \left\{ \triangle W_{-1}' Z A_N Z' \triangle Y \right\}, && (21) \end{aligned}$$

where

$$J_N = \triangle W_{-1}' Z A_N Z' \triangle W_{-1},$$

A_N is any positive definite matrix independent from θ and X_{-j} denotes the matrix X with all entries lagged j periods. The identifying equations for (21) are given by $U_N = (W'Z)\, A_N\, (Z' \triangle E) = 0$. A feasible one-step estimator $\hat{\theta}_{ABn;1}$ has

$$A_N \;=\; N^{-1}\left(Z'\left\{I_N \otimes \begin{pmatrix} 2 & 1 & & 0 & 0 \\ 1 & 2 & \ddots & & 0 \\ & \ddots & \ddots & \ddots & \\ 0 & & \ddots & 2 & 1 \\ 0 & 0 & & 1 & 2 \end{pmatrix}\right\} Z\right)^{-1} \tag{22}$$

and the optimal two-step estimator $\hat{\theta}_{ABn;2}$ has

$$A_N \;=\; V_N^{-1},$$

which can be estimated using $\hat{\theta}_{ABn;1}$. Using one of the (infeasible) estimators $\hat{\theta}_{ABn;x}$, the innovations $\triangle\varepsilon_{it}$ can then be estimated by

$$\triangle\hat{\varepsilon}_{it} \;=\; \triangle y_{it} - \hat{\alpha}_{ABn} \triangle y_{i,t-1} - \hat{\rho}_{ABn} \triangle u_{i,t-1} \tag{23}$$

and the disturbances $\triangle u_{it}$ as

$$\triangle u_{it} \;=\; \triangle y_{it} - \hat{\alpha}_{ABn} \triangle y_{i,t-1}. \tag{24}$$

Since the iterative estimator $\hat{\theta}^{(\infty)}_{ABn;x}$ based on $\hat{\theta}_{ABn;x}$ satisfies Assumption 1, following theorem follows immediately from Theorem 1.

Theorem 2. Under Assumptions 1-2, it holds that

$$\sqrt{N}\left(\hat{\theta}^{(\infty)}_{ABn;x} - \theta_0\right) \;\sim\; \mathcal{N}\left(0; \Sigma_{ABn;x}\right),$$

with

$$\Sigma_{ABn;x} \;=\; (I-M)^{-1} J^{-1}\left\{\triangle W'_{-1} Z A V A Z' \triangle W_{-1}\right\} J^{-1} (I-M')^{-1}$$

and

$$
\begin{aligned}
M &= -J^{-1}K \\
J &= \operatorname*{plim}_{N\to\infty}\left[\left.\frac{\partial U_N(\theta;\psi)}{\partial\theta'}\right|_{\substack{\theta=\theta_0\\ \psi=\psi_0}}\right] \\
&= \operatorname*{plim}_{N\to\infty}\left[\triangle W_{-1}' ZAZ' \triangle W_{-1}\right] \\
K &= \operatorname*{plim}_{N\to\infty}\left[\left.\frac{\partial U_N(\theta;\psi)}{\partial\psi'}\right|_{\substack{\theta=\theta_0\\ \psi=\psi_0}}\right] \\
&= -\rho\operatorname*{plim}_{N\to\infty}\left[\triangle W_{-1}' ZAZ' \begin{pmatrix} \triangle Y_{-2} & 0 \end{pmatrix}\right].
\end{aligned}
$$

The proposed estimator $\hat{\theta}^{(\infty)}_{ABn;x}$ extends the well-known Arellano-Bond (1991) estimator, which only allows moving average disturbances, to autoregressive disturbances. The first difference operator removes the fixed effects from y_{it}, while at the same time it preserves the autoregressive structure of the disturbances. As an alternative set of instruments one could define z_{it} as $(y_{it}, \mu_i + u_{it})'$, since the quantity $\mu_i + u_{it}$ is directly predictable from the equation in levels (19) and $\mu_i + u_{i,t-2}$ is independent from $\triangle\varepsilon_{it}$.

As an alternative, one could think of estimating (μ, α, ρ) using model (19)-(20) in levels

$$
y_{it} = \mu + \alpha y_{i,t-1} + \rho u_{i,t-1} + (\mu_i - \mu) + \varepsilon_{it},
$$

using the Arrelano-Bover (1995) instruments $(1, \triangle y_{i,t-1}, u_{i,t-1}, \ldots, \triangle y_{i2}, u_{i2})$. The disturbances u_{it} could then be estimated by

$$
\hat{u}_{it} = y_{it} - \hat{\alpha} y_{i,t-1} - (T-1)^{-1}\sum_{s=2}^{T}(y_{is} - \hat{\alpha} y_{i,s-1}). \tag{25}
$$

However, an iterative estimator based on this (unfeasible) Arrelano-Bover (1995) estimator would be inconsistent for $N \to \infty$, since $\hat{u}_{it}$ converges to $u_{it} - (T-1)^{-1}\sum_{s=2}^{T} u_{is}$.

The iterative Arrelano-Bond estimator is also applicable when other explanatory variables x_{it} are included in (15). Then the vector of instruments is expanded to $Z_{it} = \left(z_{i1}', z_{i2}', \ldots, z_{i,t-2}', x_{i1}', \ldots, x_{i,t-1}'\right)'$ in case of predetermined regressors and to $Z_{it} = \left(z_{i1}', z_{i2}', \ldots, z_{i,t-2}', x_{i1}', \ldots, x_{iT}'\right)'$ in case of strictly exogenous x_{it}.

4 Monte Carlo

In order to study the behaviour of the proposed estimator, some small Monte Carlo simulations were carried out, for which the *DGP* was specified as follows

$$\begin{aligned} y_{it} &= \alpha y_{i,t-1} + \beta_0 + \beta_1 x_{it} + \mu_i + u_{it} \\ u_{it} &= \rho u_{i,t-1} + \varepsilon_{it} \\ x_{it} &= \phi x_{i,t-1} + \xi_{it} \end{aligned}$$

for $i = 1, \ldots, N$ and $t = 1, \ldots, T$, with initial values generated by

$$\begin{aligned} y_{i0} &= (\beta_0 + \beta_1 x_{i0} + \mu_i + u_{i0}) \,/\, \left(1 - \alpha^2\right) \\ u_{i0} &= \varepsilon_{i0} / \left(1 - \rho^2\right) \\ x_{i0} &= \xi_{i0} / \left(1 - \phi^2\right), \end{aligned}$$

where $(\mu_i, \varepsilon_{i0}, \ldots, \varepsilon_{iT}, \xi_{i0}, \ldots, \xi_{iT})' \sim NID\left(0, \left[\sigma_\mu^2, I_{T+1}, I_{T+1}\right]\right)$. I focus on the behaviour of $\hat{\theta}_N^{(\infty)}$ in function of the autoregressive coefficients of both the outcome and the errors and thus I let these parameters take on the values $\alpha, \rho = 0.2,\ 0.5,\ 0.8$. The other parameters remain fixed: $\beta_0 = \beta_1 = 1$ and $\phi = 0.8$. This design is comparable to the *DGP* from Arellano and Bond (1991). Every Monte Carlo experiment consisted of 1000 replications.

In a first set of experiments, the iterated *OLS* estimator (17), with $t_0 = 3$, is studied, using the above *DGP* with $\sigma_\mu^2 = 0$. As a starting point, the inconsistent *OLS* estimate of y_{it} on $(y_{i,t-1}, 1, x_{it})$, ignoring the term $\rho u_{i,t-1}$, was used. The results are quite satisfactory and are represented in table 1, with standard errors given in parentheses. The iteration allways converged ($C = 100\%$) and the mean number of iterations till convergence is represented by R. The maximum relative errors are 2.25% for $\hat{\rho}$, 2.95% for $\hat{\alpha}$ and $(1.21, 1.43)\,\%$ for $\left(\hat{\beta}_0, \hat{\beta}_1\right)$, with no clear pattern apparent from the results. The standard errors are fairly accurately estimated, with relative errors smaller than 5%, except for $\rho = 0.8$, $\alpha = 0.2, 0.5$ where the relative error for $\hat{\sigma}_{\hat{\rho}}$ is about 15.67% and for $\hat{\sigma}_{\hat{\alpha}}$ about 7.46%.

A second set of experiments included fixed effects, with $\sigma_\mu^2 = 1$. The iterated *GMM* estimator based on (21) with A_N given by (22), i.e. the one-step *GMM* estimator, was studied. In table 2 a comparison is made with Arellano and Bonds one-step *GMM* estimator (21), for the *DGP* with $\rho = 0$. All parameter esimates estimated by both estimators are close to each other and to the true value. Except for the autoregressive parameter α, the iterative procedure does not seem overly costly in terms of variance. With respect to the one-step *GMM*, the variance of

Table 1: Monte Carlo results for iterated OLS

N, T, ϕ	100, 7, 0.8								
ρ	0.2			0.5			0.8		
α	0.2	0.5	0.8	0.2	0.5	0.8	0.2	0.5	0.8
$\hat{\rho}$	0.1969 (0.0581)	0.1955 (0.0527)	0.1956 (0.0468)	0.4949 (0.0549)	0.4937 (0.0510)	0.4951 (0.0452)	0.7846 (0.0552)	0.7904 (0.0564)	0.7920 (0.0455)
$\hat{\alpha}$	0.1988 (0.0394)	0.4992 (0.0267)	0.7990 (0.0134)	0.1992 (0.0491)	0.4993 (0.0350)	0.7997 (0.0179)	0.2059 (0.0898)	0.4972 (0.0747)	0.7988 (0.0379)
$\hat{\beta}_1$	1.0011 (0.0439)	1.0026 (0.0460)	1.0026 (0.0415)	0.9984 (0.0479)	1.0019 (0.0514)	1.0000 (0.0519)	0.9879 (0.0572)	1.0000 (0.0684)	1.0050 (0.0711)
$\hat{\beta}_0$	0.9991 (0.0734)	1.0012 (0.0755)	1.0018 (0.0723)	0.9986 (0.1065)	0.9980 (0.1171)	1.0002 (0.1066)	0.9857 (0.2508)	1.0054 (0.3112)	1.0186 (0.3191)
$\hat{\sigma}_{\hat{\rho}}$	0.0588 (0.0022)	0.0520 (0.0016)	0.0465 (0.0010)	0.0564 (0.0036)	0.0529 (0.0028)	0.0451 (0.0016)	0.0465 (0.0085)	0.0528 (0.0077)	0.0457 (0.0053)
$\hat{\sigma}_{\hat{\alpha}}$	0.0393 (0.0026)	0.0260 (0.0020)	0.0130 (0.0010)	0.0490 (0.0048)	0.0359 (0.0043)	0.0181 (0.0020)	0.0831 (0.0142)	0.0752 (0.0219)	0.0376 (0.0123)
$\hat{\sigma}_{\hat{\beta}_1}$	0.0457 (0.0022)	0.0441 (0.0024)	0.0399 (0.0024)	0.0485 (0.0026)	0.0515 (0.0030)	0.0497 (0.0031)	0.0515 (0.0054)	0.0691 (0.0080)	0.0714 (0.0092)
$\hat{\sigma}_{\hat{\beta}_0}$	0.0746 (0.0057)	0.0752 (0.0062)	0.0714 (0.0057)	0.1092 (0.0124)	0.1151 (0.0142)	0.1125 (0.0131)	0.2514 (0.0630)	0.2994 (0.1164)	0.3134 (0.1751)
C, R	100%, 7.216	100%, 7.289	100%, 7.284	100%, 13.548	100%, 13.822	100%, 14.278	100%, 38.553	100%, 41.215	100%, 41.738

Table 2: Monte Carlo results for iterated *GMM*

N, T, ϕ	100, 7, 0.8					
ρ	0					
α	0.2		0.5		0.8	
$\hat{\rho}$	-0.0201 (0.1394)	/	-0.0057 (0.1171)	/	-0.0315 (0.0880)	/
$\hat{\alpha}$	0.1974 (0.1058)	0.1832 (0.0630)	0.4581 (0.1073)	0.4625 (0.0890)	0.7601 (0.0836)	0.7620 (0.0816)
$\hat{\beta}_1$	1.0042 (0.0651)	1.0042 (0.0603)	1.0001 (0.0652)	1.0017 (0.0575)	0.9811 (0.0750)	0.9832 (0.0689)
$\hat{\beta}_0$	-0.0008 (0.0351)	-0.0008 (0.0279)	-0.0014 (0.0365)	0.0018 (0.0295)	0.0172 (0.0496)	0.0168 (0.0470)
$\hat{\sigma}_{\hat{\rho}}$	0.1699 (0.1492)	/	0.1233 (0.0657)	/	0.0850 (0.0116)	/
$\hat{\sigma}_{\hat{\alpha}}$	0.1347 (0.1020)	0.0623 (0.0061)	0.1068 (0.0528)	0.0828 (0.0161)	0.0781 (0.0300)	0.0787 (0.0160)
$\hat{\sigma}_{\hat{\beta}_1}$	0.0666 (0.0046)	0.0594 (0.0035)	0.0654 (0.0052)	0.0578 (0.0038)	0.0725 (0.0070)	0.0673 (0.0070)
$\hat{\sigma}_{\hat{\beta}_0}$	0.0361 (0.0126)	0.0272 (0.0013)	0.0360 (0.0094)	0.0278 (0.0022)	0.0458 (0.0110)	0.0451 (0.0087)
C, R	91.3%, 14.387	/	98.9%, 7.319	/	99.9%, 6.248	/

$\hat{\alpha}$ is inflated with 67.94% when $\alpha = 0.2$, but this increase diminishes to a mere 2.45% when $\alpha = 0.8$. This extra variance stems from both the extra cross-section that is lost by the iterative procedure and from the fact that an estimated regressor is used. The results for values of ρ different from zero, inform us that the small sample performance of the iterated one-step *GMM* estimator are not fantastic. More extensive simulations are called for, both to gain insight into the parameter ranges for which the behaviour becomes acceptable, and to assess its relative performance with respect to the maximum likelihood estimator.

5 Conclusion

The presented iterative estimation procedure correcting for serial correlation of the disturbances gives a promising first impression. It will be usefull in the presence of lagged dependent variables and fixed effects when the error distribution is unknown. The results in this paper, although promising, call for a more thorough investigation.

A Proofs

References

[1] Ahn, S.C., Schmidt, P. (1995), "Efficient Estimation of Models for Dynamic Panel Data", *Journal of Econometrics*, **68**(1), p.5-27.

Table 3: Monte Carlo results for iterated *GMM*

N, T, ϕ	100, 7, 0.8								
ρ	0.2			0.5			0.8		
α	0.2	0.5	0.8	0.2	0.5	0.8	0.2	0.5	0.8
$\hat{\rho}$	0.1347 (0.1489)	0.1530 (0.1360)	0.0728 (0.1079)	0.2539 (0.1721)	0.2998 (0.1509)	0.0909 (0.1315)	0.2217 (0.1585)	0.3009 (0.1506)	0.0758 (0.1292)
$\hat{\alpha}$	0.2047 (0.1126)	0.4472 (0.1186)	0.6760 (0.0985)	0.2833 (0.1299)	0.4536 (0.1183)	0.5862 (0.1019)	0.4217 (0.1321)	0.5279 (0.1319)	0.6334 (0.0902)
$\hat{\beta}_1$	1.0100 (0.0638)	1.0160 (0.0694)	0.9769 (0.0673)	1.0269 (0.0619)	1.0480 (0.0631)	0.9781 (0.0685)	1.0403 (0.0619)	1.0708 (0.0619)	1.0016 (0.0635)
$\hat{\beta}_0$	-0.0013 (0.0411)	-0.0042 (0.0421)	0.0323 (0.0511)	0.0007 (0.0501)	-0.0177 (0.0531)	0.0594 (0.0610)	-0.0012 (0.0494)	-0.0211 (0.0562)	0.0403 (0.0625)
$\hat{\sigma}_{\hat{\rho}}$	0.3015 (0.6407)	0.2108 (0.3401)	0.1057 (0.0390)	0.7328 (3.3844)	0.7258 (7.7189)	0.1231 (0.0760)	0.4443 (2.4440)	3.8574 (101.6767)	0.1070 (0.0390)
$\hat{\sigma}_{\hat{\alpha}}$	0.2033 (0.3667)	0.1701 (0.2978)	0.1046 (0.1503)	0.4623 (2.1165)	0.7870 (12.0818)	0.1051 (0.2376)	0.3004 (1.5563)	4.5034 (119.5385)	0.0712 (0.0896)
$\hat{\sigma}_{\hat{\beta}_1}$	0.0674 (0.0208)	0.0671 (0.0166)	0.0688 (0.0228)	0.0736 (0.0664)	0.1278 (1.2167)	0.0604 (0.0294)	0.0686 (0.0588)	0.4174 (9.9589)	0.0533 (0.0059)
$\hat{\sigma}_{\hat{\beta}_0}$	0.0502 (0.0797)	0.0502 (0.0734)	0.0504 (0.0366)	0.0892 (0.2336)	0.2720 (4.4953)	0.0450 (0.0515)	0.0741 (0.3214)	0.7731 (19.6170)	0.0347 (0.0136)
C, R	90.8%, 14.863	95.1%, 8.639	99.2%, 7.850	70.4%, 16.455	82.2%, 13.238	96.1%, 8.734	85.2%, 12.419	85.6%, 14.688	98.5%, 8.021

[2] Ahn, S.C., Schmidt, P. (1997), "Efficient Estimation of Dynamic Panel Data Models: Alternative Assumptions and Simplified Estimation", *Journal of Econometrics*, **76**(1-2), p.309-321.

[3] Aitkin, M., Alfò, M. (2003), "Longitudinal Analysis of Repeated Binary Data using Autoregressive and Random Effect Modelling", *Statistical Modelling*, **3**(4), p.291-303.

[4] Anderson, T.W., Hsiao, C. (1981), "Estimation of Dynamic Models with Error Components", *Journal of the American Statistical Association*, 76(375), p.598-606.

[5] Arellano, M., Bond, S. (1991), "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations", *Review of Economic Studies*, **58**(2), p.277-297.

[6] Arellano, M., Bover, O. (1995), "Another Look at the Instrumental Variable Estimation of Error-Components Models", *Journal of Econometrics*, **68**(1), p.29-51.

[7] Arslan, O., Constable, P.D.L., Kent, J.I. (1993), "Domains of Convergence for the *EM* Algorithm: A Cautionary Tale in a Location Estimation Problem", *Statistics and Computing*, **3**, p.103-108.

[8] Baltagi, B.H., Li, Q. (1991), "A Transformation That Will Circumvent the Problem of Autocorrelation in an Error Component Model", *Journal of Econometrics*, **48**, p.385-393.

[9] Baltagi, B.H., Li, Q. (1995), "Testing $AR(1)$ against $MA(1)$ Disturbances in an Error Component Model", *Journal of Econometrics*, **68**(1), p.133-151.

[10] Bhargava, A., Franzini, L., Narendranathan, W. (1982), "Serial Correlation and the Fixed Effects Model", *Review of Economic Studies*, **49**(4), p.533-549.

[11] Chaganty, N.R. (1997), "An alternative approach to the analysis of longitudinal data via generalized estimating equations", *Journal of Statistical Planning and Inference*, **38**(3), p.371-380.

[12] Cox, D.R., Hinkley, D.V. (1974), *Theoretical Statistics*, Chapman and Hall, London.

[13] Cox, D.R., Snell, E.J. (1968), "A General Definition of Residuals", *Journal of the Royal Statistical Society B*, **30**(2), p.248-275.

[14] Crépon, B., Kramarz, F., Trognon, A. (1997), "Parameters of Interest, Nuisance Parameters and Orthogonality Conditions—An Application to Autoregressive Error Component Models", *Journal of Econometrics*, **82**(1), p.135-156.

[15] Dempster, A.P., Laird, N.M., Rubin, D.B. (1977), "Maximum Likelihood from Incomplete Data via the *EM* Algorithm", *Journal of the Royal Statistical Society, Series B (Methodological)*, **39**(1), p.1-38.

[16] Dominitz, J., Sherman, R.P. (2005), "Some Convergence Theory for Iterative Estimation Procedures with an Application to Semiparametric Estimation", *Econometric Theory*, **21**(4), p.838-863.

[17] Doran, H.E., Schmidt, P. (2006), "*GMM* estimators with improved finite sample properties using principal components of the weighting matrix, with an application to the dynamic panel data model", *Journal of Econometrics*, **133**(1), p.387-409.

[18] Elashoff, M., Ryan, L. (2004), "An *EM* Algorithm for Estimating Equations", *Journal of Computational and Graphical Statistics*, **13**(1), p.48-65.

[19] Gouriéroux, C., Monfort, A., Trognon, A. (1985), "A General Approach to Serial Correlation", *Econometric Theory*, **1**(3), p.315-340.

[20] Greene, W.H. (1997), *Econometric Analysis*, Prentice-Hall, London.

[21] Hamilton, J. D. (1994), *Time Series Analysis*, Princeton University Press.

[22] Hansen, L.P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators", *Econometrica*, **50**(4), p.1029-1054.

[23] Horn, R.A., Johnson, C.R. (1994), *Matrix Analysis*, Cambridge University Press.

[24] Hsiao, C., Pesaran M.H., Tahmiscioglu, A.K. (2002), "Maximum Likelihood Estimation of Fixed Effects Dynamic Panel Data Models Covering Short Time Periods", *Journal of Econometrics*, **109**, p.107-150.

[25] Kapetanios, G. (2003), "A Note on an Iterative Least-Squares Estimation Method for *ARMA* and *VARMA* Models", *Economics Letters*, **79**(3), p.305-312.

[26] Koreisha, S., Pukkila, T.(1990), "Linear Methods for Estimating Autoregressive Moving Average and Regression Models with Serial Correlation", *Communications in Statistics - Simulation and Computation*, **19**, p.71-102.

[27] Liang, K.-Y., Zeger, S.L. (1986), "Longitudinal Data Analysis Using Generalized Linear Models", *Biometrika*, **73**(1), p.13-22.

[28] Lillard, L.A., Willis, R.J. (1978), "Dynamic Aspects of Earning Mobility", *Econometrica*, **46**(5), p.985-1012.

[29] McCullagh, P., Nelder, J. A. (1999), *Generalized Linear Models, 2nd ed.*, London : Chapman and Hall.

[30] Murphy, K.M., Topel, R.H. (1985), "Estimation and Inference in Two-Step Econometric Models", *Journal of Business and Economic Statistics*, **3**(4), p.370-379.

[31] Oakes, D. (1999), "Direct Calculation of the Information Matrix via the *EM* Algorithm", *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, **61**(2), p.479-482.

[32] Parke, W.R. (1986), "Pseudo Maximum Likelihood Estimation: The Asymptotic Distribution", *Annals of Statistics*, **14**(1), p.355-357.

[33] Pastorello, S., Patilea, V., Renault, E. (2003), "Iterative and Recursive Estimation in Structural Nonadaptive Models", *Journal of Business and Economic Statistics*, **21** (4), p.449-482.

[34] Pierce, D.A. (1982), "The Asymptotic Effect of Substituting Estimators for Parameters in Certain Types of Statistics", *Annals of Statistics*, **10**(2), p.475-478.

[35] Randles, R.H. (1982), "On the Asymptotic Normality of Statistics with Estimated Parameters", *Annals of Statistics*, **10**(2), p.462-474.

[36] Searle, S.R., Casella, G., McCulloch, C.E. (1992), *Variance Components*, Wiley Series in Probability and Mathematical Statistics, Applied Probability and Statistics, Wiley, N.Y.